# Calculating MODFLOW Analytical Sensitivities Using ADIFOR for Effective and Efficient Estimation of Uncertainties

Amir Gamliel[1],
U.S. Army Corps of Engineers, Interagency modeling Center,
3301 Gun Club Road, West Palm Beach, Florida 33406.


Mike Fagan[2],
Department of Computational and Applied Mathematics,
Rice University, 6100 Main St. - MS 134
Houston, TX 77005-1892

Maged Hussein
U.S. Army Corps of Engineers, Interagency modeling Center,
3301 Gun Club Road, West Palm Beach, Florida 33406.

## Abstract

A modified MODFLOW code was developed to calculate the analytical sensitivities of both the field parameters and the operational parameters using the automatic differentiation system ADIFOR 3.0. The analytical sensitivities were then used as input parameters for PEST to calculate estimates for the uncertainties. In addition, the manuscript details a comparison between the analytical sensitivities and numerical estimates of the sensitivities as calculated by using PEST, and presents a comparison between the estimates of the uncertainties based on analytical sensitivities and numerically derived sensitivities. Analytical calculations of the sensitivities are known to be more accurate than numerical estimates, usually takes less time to compute, and the results of analytical computations are independent of numerical assumptions such as the magnitude of the perturbation.

# Introduction

Edward Lorenz conducted research at MIT by simulating weather patterns on the university's huge (in physical dimensions) computer about forty years ago. One winter day in 1961 he decided to make a shortcut and punched a value of 0.506 on one of the input data cards instead of 0.506127 he used in previous simulations. Something unexpected happened. He noticed that the new simulation results dramatically diverged from the previous one. At the beginning, Lorenz thought that the MIT computer had a malfunction. Then, after some research, he discovered that the difference in the simulated weather pattern is due to rounding up one of the initial condition parameters by one thousandth. That slight change in the initial conditions produced a huge difference in the result.

Lorenz's experiment discovered the phenomenon of sensitive dependence on initial conditions, or the butterfly effect: the change brought by a butterfly moving its wings in Shanghai on the weather in San Francisco. Since Lorenz's experiment, scientists and engineers have studied sensitivities of models with respect to boundary conditions, management decisions, resource allocations, and discretizations of numerical models.

Uncertainties in modeling results are due to various factors, including the assumptions made in the development of conceptual mathematical models used to describe the physical phenomena to be modeled, the assumptions in numerical algorithms developers use to convert mathematical models into numerical models, temporal and spatial discretization, and errors in field measurements used for model calibration or validation. The goal of an uncertainty analysis task is to quantify uncertainty in the outcome of the model under consideration.

Both sensitivity and uncertainty analysis require accurate estimates of the derivatives of the model output with respect to the uncertain parameters. This manuscript demonstrates the use of an automatic differentiation utility to accurately and efficiently estimate analytical derivatives of model output with respect to input parameters. These derivatives can be used for automatic model calibration, sensitivity analysis, and uncertainty analysis. The Modular Three-Dimensional Finite-Difference Ground-Water Flow Model (MODFLOW) was used in this paper to demonstrate how the methodology can be utilized in groundwater applications.
A new MODFLOW code was developed to analytically compute the derivatives of the dependent variables with respect to the independent variables. The manuscript compares the resulting analytical derivatives with the derivatives computed by the numerical method using perturbation. The derivatives are used as input data to compute uncertainties more efficiently.

# Sensitivity Analysis

Engineers have extensively relied on computer simulations for decision support. Model output allows engineers to study and predict outcomes under various management scenarios and alternatives. However, model predictions can be misleading if the underlying uncertainty is not quantified. In the case of ground water, parameters used in the models can vary a few

orders of magnitude rendering model predictions uncertain. Sensitivity analysis can be a valuable tool for engineers to examine the importance of parameter uncertainty during both model calibration and evaluation of model predictions.

Scientists and engineers usually perform sensitivity analysis with respect to four types of parameters: field parameters, parameters related to boundary conditions, decision parameters, and parameters related to the numerical algorithm. Field parameters include a wide range of parameters. Examples of field parameters for groundwater modeling include viscosity, transmissivity, conductivity, dispersivity, density, heat capacity, chemical absorption, and reaction rates. State variables include pressure, velocity, and concentration. Boundary conditions parameters are state variables that are specified along the boundary of the area of interest and include head, velocity, and concentration. Decision parameters include allocation of resources to maximize or minimize system output, subject to given constrains. Examples of decision parameters in groundwater modeling include pumping and injection rates, drain elevation, drain conductance, etc. Optimization of such allocation results in maximum utilization and preservation of the aquifer. This work presents, as an example, sensitivity of an independent variable (pressure head in an aquifer system) with respect to two field parameters (transmissivity and vertical conductivity), one boundary condition parameter (specified head along a river), and one decision parameter (pumping rate of a well system).

To date, there are two main techniques available to conduct sensitivity analysis: the analytical approach and the numerical approach. The analytical approach calculates the value of the analytical derivative of a model output with respect to model parameters. The numerical perturbation approach is based on calculating the derivatives by incrementally varying the parameter values using small perturbations. Some of the available codes for conducting sensitivity analysis by the perturbation method include the Model-Independent Parameter Estimation code (PEST) (Doherty, 2002) and the Computer Code for Universal Inverse Modeling (UCODE) (Poeter and Hill, 1998). For the effort described herein, we elected to utilize PEST for sensitivity analysis using the numerical approach. Sensitivity analysis using the analytical approach is preferable over numerical differentiation methods because the resulting analytical derivatives are more accurate and efficient in terms of computer resources needed to perform the calculation. In the current work, analytical derivatives were obtained by utilizing an innovative automatic differentiation technique.

*Automatic Differentiation*
Code development for analytical derivatives is very difficult and tedious for complex numerical models which involve thousands of lines of computer code. Examples of manually developed analytical derivatives in computer codes include MODFLOW-PES processes (Harbaugh et al., 2000) and the code developed by Manganelli et al. (2002). A more efficient alternative is Automatic Differentiation (AD). In the current work the innovative approach of AD was found to be of great value for developing analytical derivatives. AD is a technique for augmenting modeling codes with partial derivative computations. It exploits the fact that every code executes a sequence of elementary arithmetic operations. Martins et al., (2000) and Griewank (2000) discuss the mathematical and computational techniques used for general purpose AD computer programs. The

Automatic Differentiation of Fortran (ADIFOR 3.0) (Fagan and Carle, 2001) was selected for use in this paper. This development tool automatically augments the original code with new code for calculating partial derivatives of selected output variables with respect to a set of input parameters. The technique and the resulting computer program are so flexible that the user has full control of the number and size of the dependent and independent variables. From the user's point of view, ADIFOR functions much like a compiler: the user designates the source code to be differentiated, then ADIFOR reads and analyzes it, and finally, ADIFOR writes out a new code that computes the partial derivatives in addition to the original code.

In this effort we utilized ADIFOR on the popular groundwater flow modeling code MODFLOW. ADIFOR allowed us to calculate a wide range of partial derivatives of state variables including head, discharge, and budget terms among others with respect to any input parameter. Unlike the approach used by Harbaugh et al. (2000) in MODFLOW-PES where analytical derivatives with respect to some of the boundary conditions and some of the parameters could not be calculated, ADIFOR allowed us to examine the parameter structure by allowing the model to calculate the spatial variation of derivatives in space and time. To demonstrate the utility of the approach a series of simple hypothetical tests were conducted and compared to the numerical perturbation method.

When using ADIFOR to develop the analytical solution of the derivatives, the user specifies for which dependent variables to calculate the derivatives and with respect to which independent variables. Thus, we used the opportunity to develop a modified MODFLOW-88 code to compute the derivatives of the heads with respect to all of the independent variables, including field parameters (vertical conductance, conductivities along columns, conductivities along rows, and primary and secondary storage coefficients), boundary conditions (head prescribed in boundary cells (general head boundary condition), recharge rates, maximum evapotranspiration rate, evaporation extinction depth, river water heads, river bed conductance, drain elevations, drain conductance, well pumping rates, reservoir water heads, and reservoir conductance), and parameters of the numerical algorithm (grid spacing along rows, grid spacing along columns, and thickness of layers). It should be noted that computing analytical derivatives of the independent variables with respect to grid spacing is straight forward; it creates a challenge to compute such a derivative with the perturbation method. This is because employing the perturbation method to compute derivatives with respect to grid spacing requires distortion of the grid.

The authors performed a number of tests viz. nonlinear problems, complex combination of boundary conditions, and ground water/surface water interaction problems. In some of these tests, numerical derivatives deviate considerably when compared to the analytical derivatives obtained through automatic differentiation. The issue of reliability of numerical derivatives has been a recognized limitation in inverse modeling, sensitivity analysis, and uncertainty analysis (see e.g. Doherty 2002, Hill 1998, Poeter and Hill1998). For illustration purposes the next section details a groundwater simulation example having a simple homogeneous domain, where the transmissivity or vertical conductance could be lumped into one parameter. Overall, calculations of analytical derivatives were noticeably faster than

calculation of numerical derivatives by PEST. The cost saving in run time was considerably higher for simulations of non-homogenous conditions.

*Example*

Figure 1 illustrates the first test problem, which consists of two isotropic and homogenous confined aquifers that are separated by a leaky confining unit. The horizontal hydraulic conductivity of the two aquifers is set at 10 ft/day while the vertical conductivity is set at 1 ft/day. The upper aquifer receives recharge of 0.004 ft/day. Both the upper and lower aquifers drain into a river whose bed resistance is 1000 ft$^2$/day and having a water elevation of 75 ft. The bottom aquifer contains two wells that pump at a rate of 35000 ft$^3$/day. The problem was discretized into 2 layers and 10 by 15 square cells of 500 ft side length. The quasi 3-dimensional approach of MODFLOW (Mcdonald and Harbaugh 1988) was used to represent leakage between the two aquifers through the semi-confining unit. Figure 2 shows the resulting head distribution in the upper and lower aquifers.
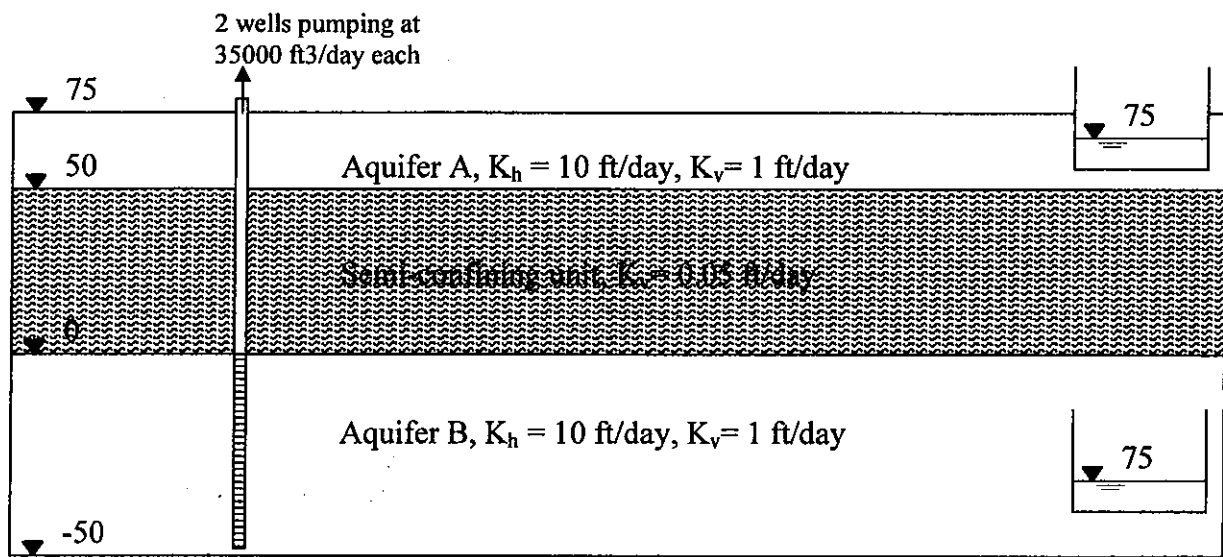


**Figure 1. Hypothetical test problem showing the use of analytical derivatives approach.**

Under the assumption that parameter uncertainty is the dominant source of uncertainty in the above problem (hence ignoring any uncertainty in the conceptual model), the following parameters can be identified as uncertain; transmissivities of the two aquifers, vertical conductance, and conductance of the river bed of the confining unit as field parameters; water level in the river and recharge as boundary conditions; pumping rate as a decision parameter; and grid spacing as the analyst decision for computational grid spacing. For demonstration purposes, this work elected to perform sensitivity and uncertainty analyses with respect to transmissivity, vertical conductivity, water level at the river, and pumping rate, with aquifer pressure heads as the dependent variable. Except for the dependency of the pressure head with respect to the pumping rate, the problem is linear and the values of the derivatives calculated accurately by numerical differentiation should be relatively close to the values calculated analytically.
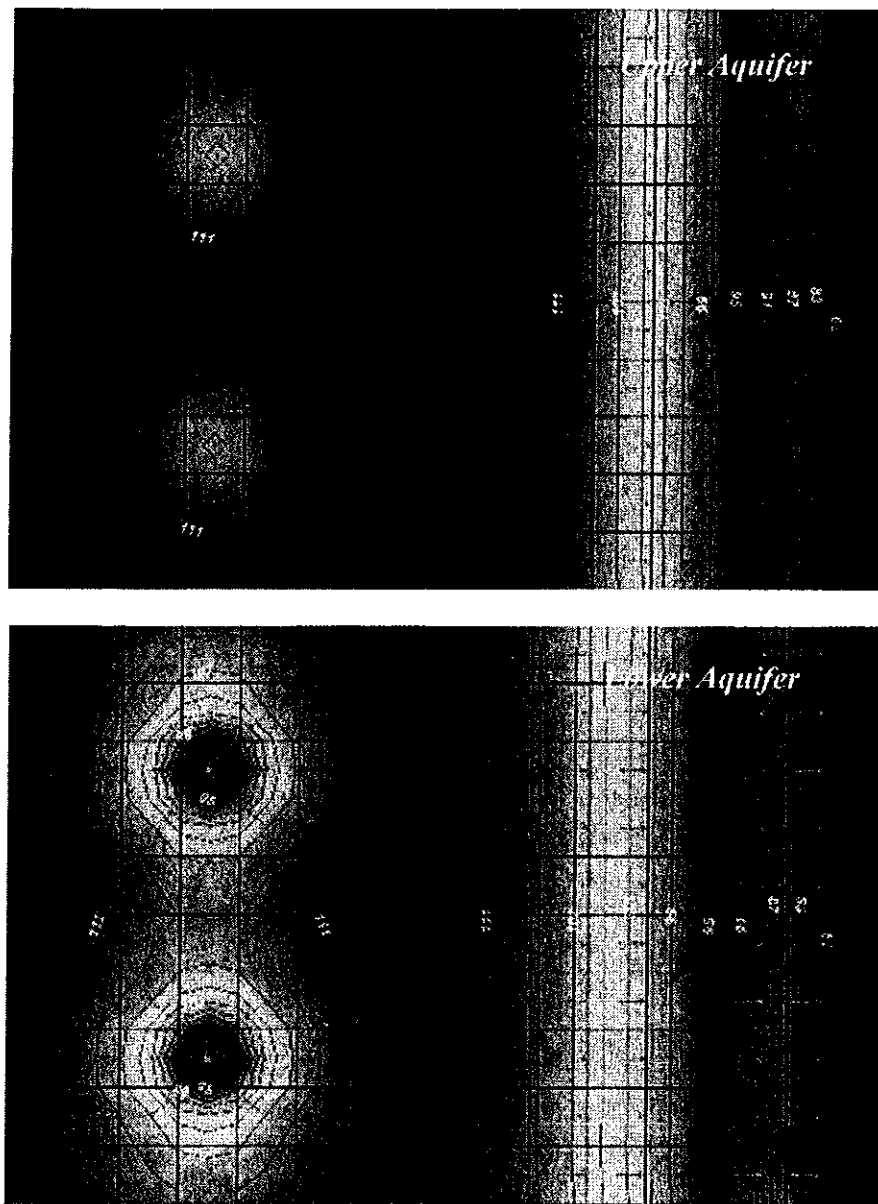
**Figure 2 Head distribution and flow field in the upper and lower aquifers.**

Despite the fact that the test problem is simple and predominantly linear, numerical differentiation produced considerably different derivatives for different differentiation methods and perturbation intervals. Figure 3 depicts the result of both analytical and numerical calculations of the derivatives of head with respect to transmissivity, vertical conductance, and pumping rate. The figure also illustrates the changes in the numerical derivatives with the change in the perturbations (increments) used for the numerical calculation. The authors change the example problem for the calculation of the derivatives of aquifer head with respect to river stage by fixing the boundary head on the left side of Figure 1 to 100 ft. Otherwise, the derivatives of aquifer heads with respect to river stage are unity
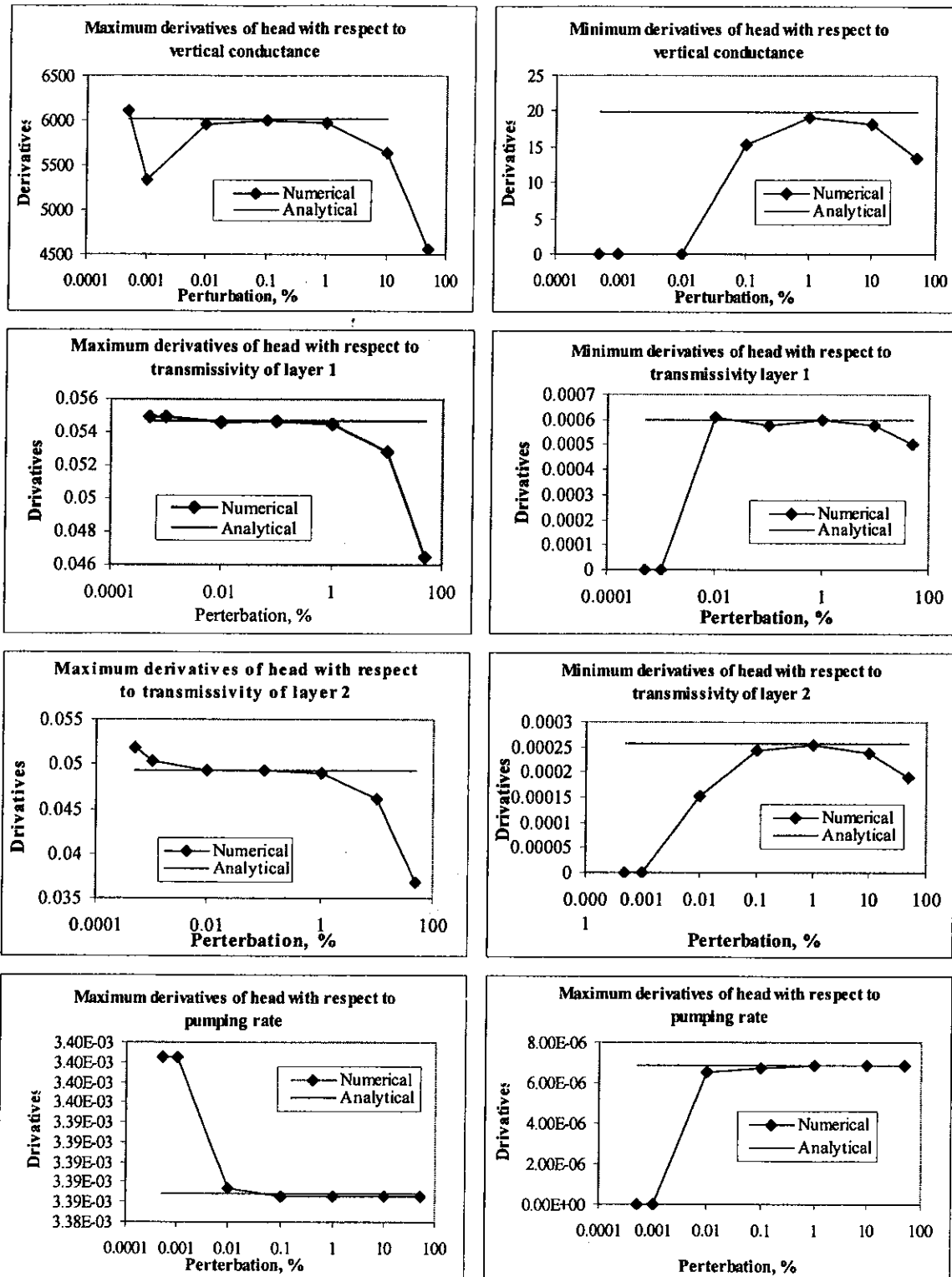
**Figure 3 Comparison between analytical numerical derivatives using different perturbations (increments) showing the reliability of the numerical derivative calculations in a simple problem.**
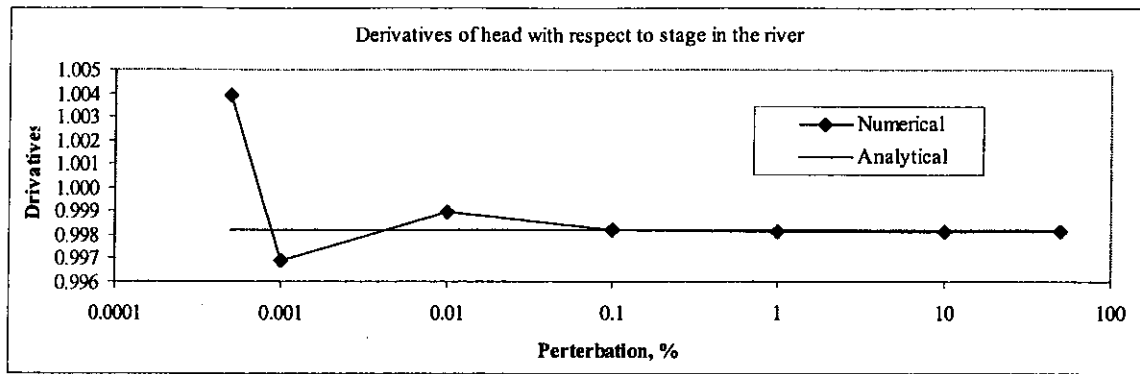
**Figure 4 Comparison between analytical and numerical derivatives using different perturbations (increments) for boundary conditions.**

over the entire domain. Figure 4 depicts the changes for the derivatives of aquifer heads with respect to river head for both the analytical and numerical calculations. Both Figures 3 and 4 exhibit that the numerical derivatives vary considerably over the range of chosen perturbations. However, the values of the numerical derivatives are close to the analytical derivatives for a given range of perturbation.

Calculations of the analytical derivatives were noticeably faster than the calculations of the numerical derivatives by PEST. The cost saving in run time was considerably higher when the above problem was slightly modified by considering a non-homogenous hydraulic conductivity field and requesting PEST and MODFLOW (with the analytical derivatives) to estimate the derivatives of each head with respect to the transmissivity of each grid cell. The analytical derivatives approach provided a considerable cost saving over PEST with respect to runtime, where PEST required two orders of magnitude time than MODFLOW with analytical derivatives.

## Uncertainty Analysis

Users of computer models employ uncertainty analysis mainly for the following purposes: examination of how much confidence one should have in the model prediction; study of the effect of errors in field parameters, used as input data, on model calibration and verification; investigation of the propagation of errors in the input data on the predicted results; improvement of model calibration with availability of additional field data; and decreasing uncertainty of model results and obtain further calibration of the input parameters by minimization of uncertainty-based error functions (inverse problem).

The uncertainty analysis methods can be grouped in four categories: respond surface or statistical methods, fussy logic methods, cross validation methods, and minimization of a generalized Baysian loss function methods. For the respond surface method, the analyst first chooses a small subset of the system parameters using their "engineering judgment." In the second step the analyst selects a perturbation pattern for the selected parameters from experimental design theories. In the third step the analyst uses the perturbation pattern to perform multiple code runs that provide new response (output) values. In the fourth step the analyst employs simple functional forms to fit the response data to a multidimensional

response surface based on a pre-determined probabilistic distribution that simplifies the original complexity of the model in question. In the fifth step the analyst utilizes sophisticated statistical algorithms (e.g., moment matching, Monte Carlo) to estimate statistical properties and uncertainty distributions of the responses.

For the minimization of a generalized Bayesian loss function the analyst solves an optimization problem (minimization of a constrained Lagrangian function) of the residual vector multiplies with some form of the covariance matrix. The optimization problem consistently combines field measurements with model outputs to simultaneously obtain best estimates for model parameters and reduce uncertainties in model outputs. This method considers all model parameters; thus, it guarantees that no important effects are overlooked by generating a full set of sensitivities. A full set means that the sensitivities with respect to all parameters are computed, without making an *a-priori* judgment as to which one is important.

To demonstrate the use of analytical derivatives in uncertainty analysis, the model described in Figure 1 is used to predict total discharge to the river from both aquifers. The discharge is estimated at 8000 ft3/day by MODFLOW. The modified MODFLOW is used to calculate both the derivatives of head at observation points and the total discharge to the river. The derivatives calculated using the modified MODFLOW are then fed to PEST to calculate the uncertainty in predicted discharge to the river. PEST calculates upper and lower bounds of the predicted discharge. The upper and lower bounds of the discharge are ~85500 ft3/day and ~76300 ft3/day respectively. These bounds are estimated as the upper and lower critical points at a predefined value of the objective function (Doherty 2002). The same uncertainty calculations were performed using PEST numerical derivatives. Both methods produced similar results. The advantage of using analytical derivatives was limited to the cost saving in runtime and accuracy. For example, in the case of numerical derivatives, MODFLOW needed to be modified to write the head and flux values with greater precision in order for PEST to converge to the predefined value of the objective function. This issue was irrelevant in the case of analytical derivatives supplied by the modified version of MODFLOW. The analytical derivatives are calculated accurately by MODFLOW and provided to PEST directly and the solution converged faster to the predefined value of the objective function.

## Conclusion

The presented work illustrates the great promise offered by automatic differentiation for analytically computing derivatives. The results of this work demonstrated that analytical calculations are more accurate, take less time to compute, and their values are not functions of the size of perturbations the analyst has chosen to use, or the method of differentiation. As a result, automatic differentiation provide great benefits for both sensitivity and uncertainty analysis especially in the case of problems with complex parameter structure, spatial and temporal variations in parameters, and highly nonlinear problems.

# References

Cooley, R. L., and R.L. Naff, 1990, Regression modeling of ground-water flow: U.S. Geological Survey Techniques in Water Resources Investigations, book 3, chap, B4, 232#p.

Doherty, J. 2002. PEST; Model-Independent Parameter Estimation; User Manual: 5th edition. Watermark Numerical Computing, Brisbane, Australia.

Griewank, A., 2000. "\Evaluating derivatives: principles and techniques of algorithmic differentiation, SIAM, Philadelphia, 369 P.

Fagan, M., and A. Carle, 2001. Adifor 3.0: Design and Implementation., Rice University Technical Report, 2001.

Harbaugh A.W., Banta E.R., Hill M.C., and McDonald, M.G., 2000. MODFLOW-2000, U.S. Geological Survey Modular Ground-Water Model—User Guide to Modularization Concepts and the Ground-Water Flow Process. U.S. Geological Survey, Open-File Report 00-92.

Hill, M.C., 1992, A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional, ground-water flow model using nonlinear regression, U.S. Geological Survey Open-File Report 91-484, 358 p.

Hill, M.C., 1998. Methods and guidelines for effective model calibration: U.S. Geological Survey, Water-Resources Investigations Report 98-4005.

Mcdonald, M. G. & A. W. Harbaugh. 1988. A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, U.S. Geological Survey Techniques of Water Resources Investigations, BSook 6, Chapter A!, 586 p.

Manganalli, S., Valadino Ceci, and Walter Vecchiato, 2002. Sensitivity Analysis of Volatility: A New Tool for Risk Management, European Central Bank, Working Papar No. 194.

Martins J.R.R.A., Ilan Kroo, and Juan J. Alonso, 2000. An Automated Method for Sensitivity Analysis Using Complex Variables, AIAA-2000-0689.

Park, C.K., K.J. Park, K.S. Lee, D.S. Bae, and H.J. Cho, . Development of an Efficient Algorithm for Dynamic and Design Sensitivity Analysis of Railway Vehicle Systems.

Poeter, E.P. and Hill, M.C., 1998, DOCUMENTATION OF UCODE, A Computer Code for Universal Inverse Modeling. U.S. Geological Survey, Water-Resources Investigations Report 98-4080.

## Authors Biography

Dr. Gamliel has been developing subsurface flow and transport simulation software and site specific non-aqueous contaminant transport, groundwater, and surface water models for the last 25 years. He conducted successful environmental site investigation and design cost-effective remediation plans, designed stormwater management plans, analyzed stream flows and floods, and designed water reservoirs and dams.

Dr. Mike Fagan received his B.A. in mathematics and electrical engineering, M.S. in electrical engineering and computation and applied mathematics, and Ph.D. in Computer Science, all from Rice University. He joined the Rice University Research Staff in 1991. His Research includes automatic differentiation and design and implementation of high level programming languages.

Dr. Maged Hussein has been developing and implementing water resources models for the last 18 years for the government, academia and privet sectors. He worked as a senior engineer/project manager for BEM systems, HydroGeoLogic and most recently for the US Army Corps of Engineer.